Preliminary work on the Statistical Emulation of a Regional Climate Model

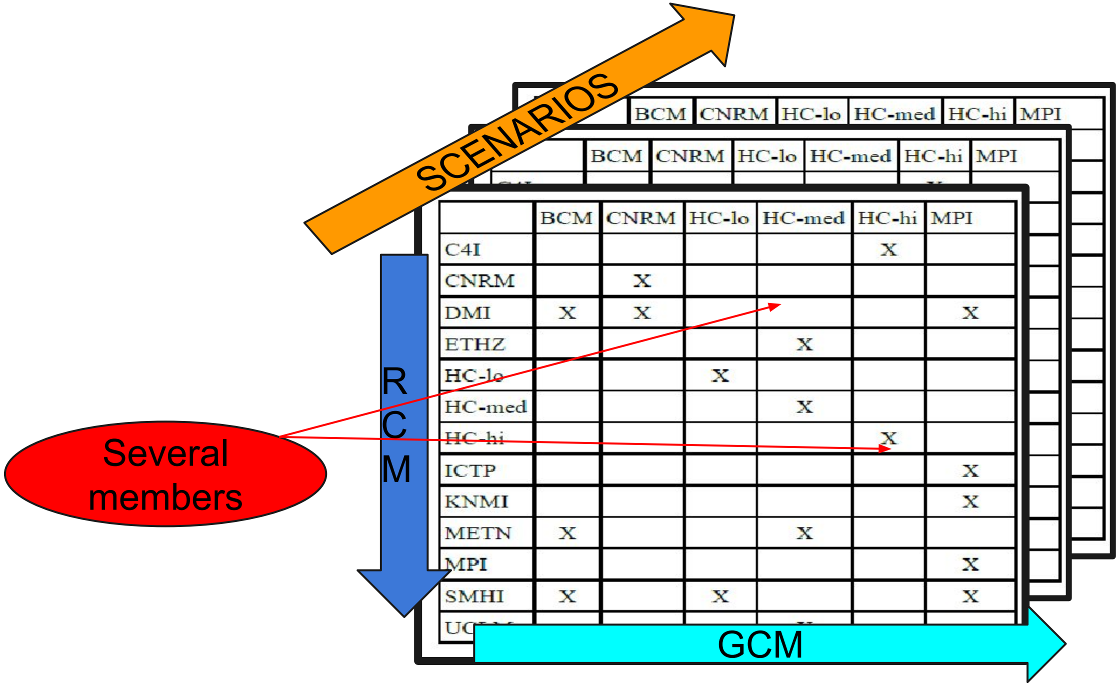
Antoine Doury, Samuel Somot, Aurélien Ribes, Lola Corre

One of the objectives of the FPS-Convection :

Is it possible to replace costly convection permitting experiments with physically defensible statistical downscaling approaches such as "convection emulators" that mimic CPMs and are fed by output of conventional scale RCMs?

Why do we need statistical emulator ?

Why do we need statistical emulator ?

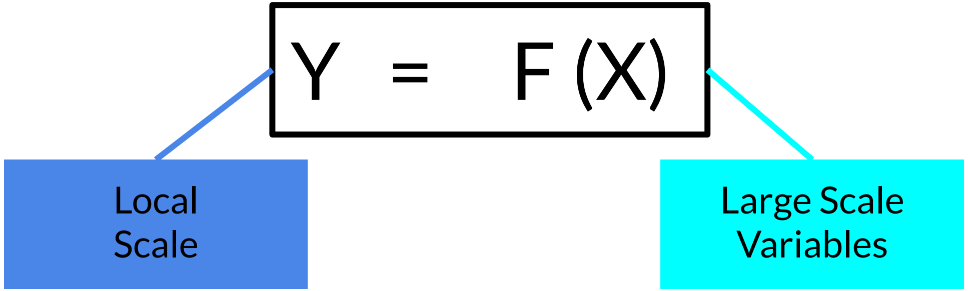


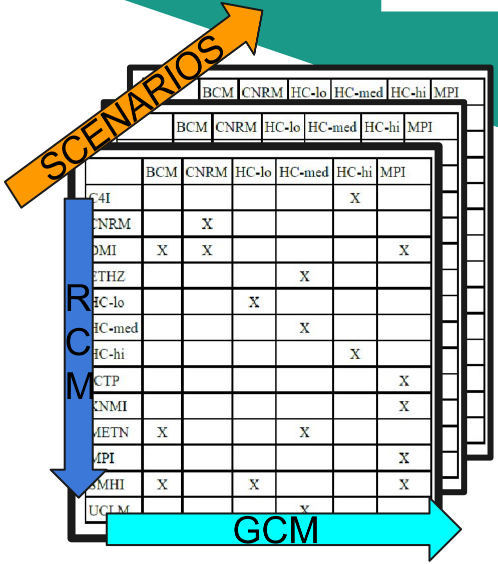
	BCM	CNRM	HC-lo	HC-med	HC-hi	MPI
C4I					X	
CNRM		X				
DMI	X	X				X
ETHZ				X		
HC-lo			X			
HC-med				X		
HC-hi					X	
ICTP						X
KNMI						X
METN	X			X		
MPI						X
SMHI	X		X			X

CP-RCM

Emulation of Regional Climate Models

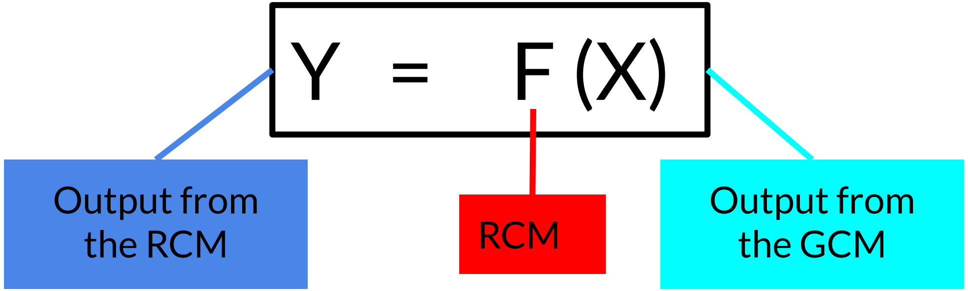
Idea : Combine both downscaling approaches to fill up the **[SCENARIO x GCM x RCM]** matrix to cover the full range of uncertainty at a reasonable cost.

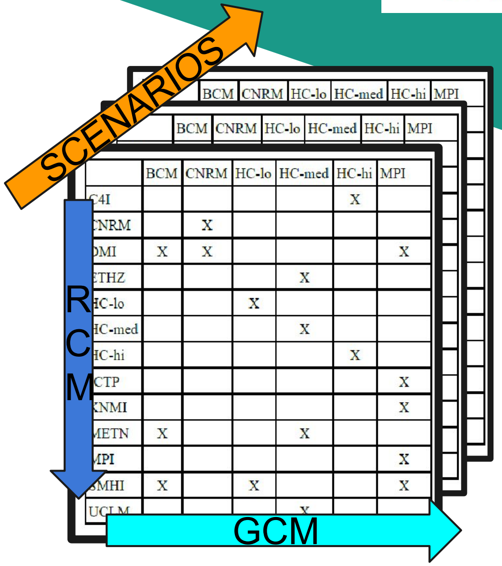




Emulation of Regional Climate Models

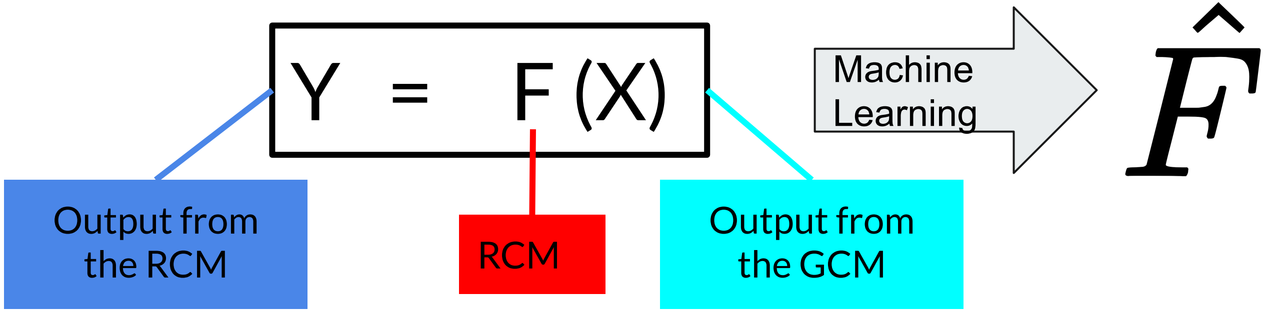
Idea : Combine both downscaling approaches to fill up the **[SCENARIO x GCM x RCM]** matrix to cover the full range of uncertainty at a reasonable cost.

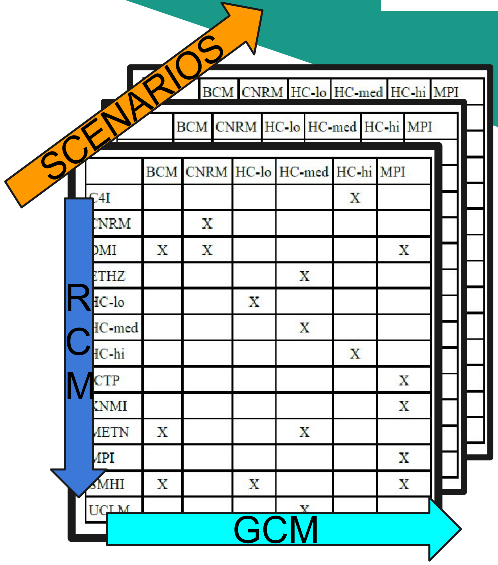




Emulation of Regional Climate Models

Idea : Combine both downscaling approaches to fill up the **[SCENARIO x GCM x RCM]** matrix to cover the full range of uncertainty at a reasonable cost.





- Advantages :
  - Learn the future relationship (no question of transferability) and on the whole grid of the RCM.
  - Computationally cheaper than RCMs.
- Limitations :
  - Strongly dependant on the quality of RCM
  - 1 emulator by RCM

Emulation of Regional Climate Models

Idea : Combine both downscaling approaches to fill up the **[SCENARIO x GCM x RCM]** matrix to cover the full range of uncertainty at a reasonable cost.

$$Y = F(X)$$

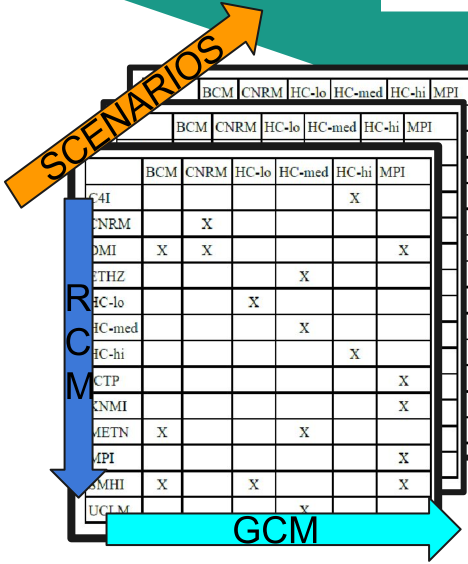
Output from the RCM

RCM

Output from the GCM

Machine Learning

$$\hat{F}$$



To start : 1 variable : TAS
1 grid point : Montpellier

XGBOOST

- Famous machine learning algorithm (classification, regression) .
- Could be a Statistical Downscaling method.
- Based on :

Regression Trees

Boosting

XGBOOST

Regression Trees

$$(\,X_i\ ,\ Y_i\,)_{i\in\{1,..,N\}}$$

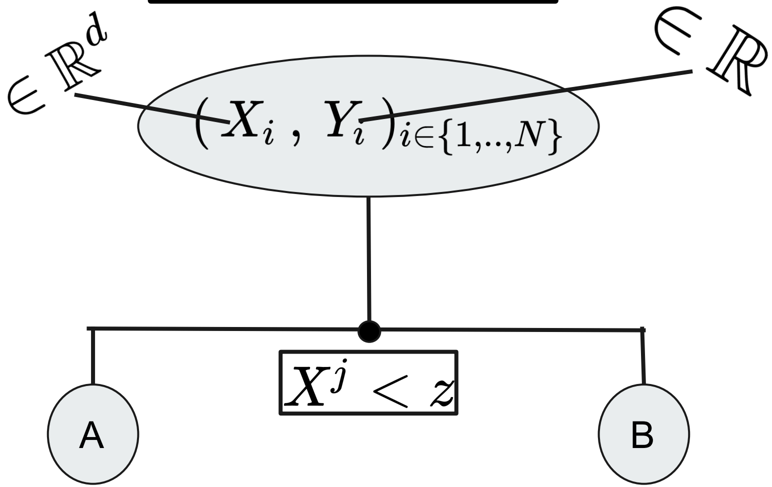
XGBOOST

Regression Trees

$$\left(\overset{\in \mathbb{R}^d}{X_i}\ ,\ \overset{\in \mathbb{R}}{Y_i}\right)_{i\in\{1,..,N\}}$$

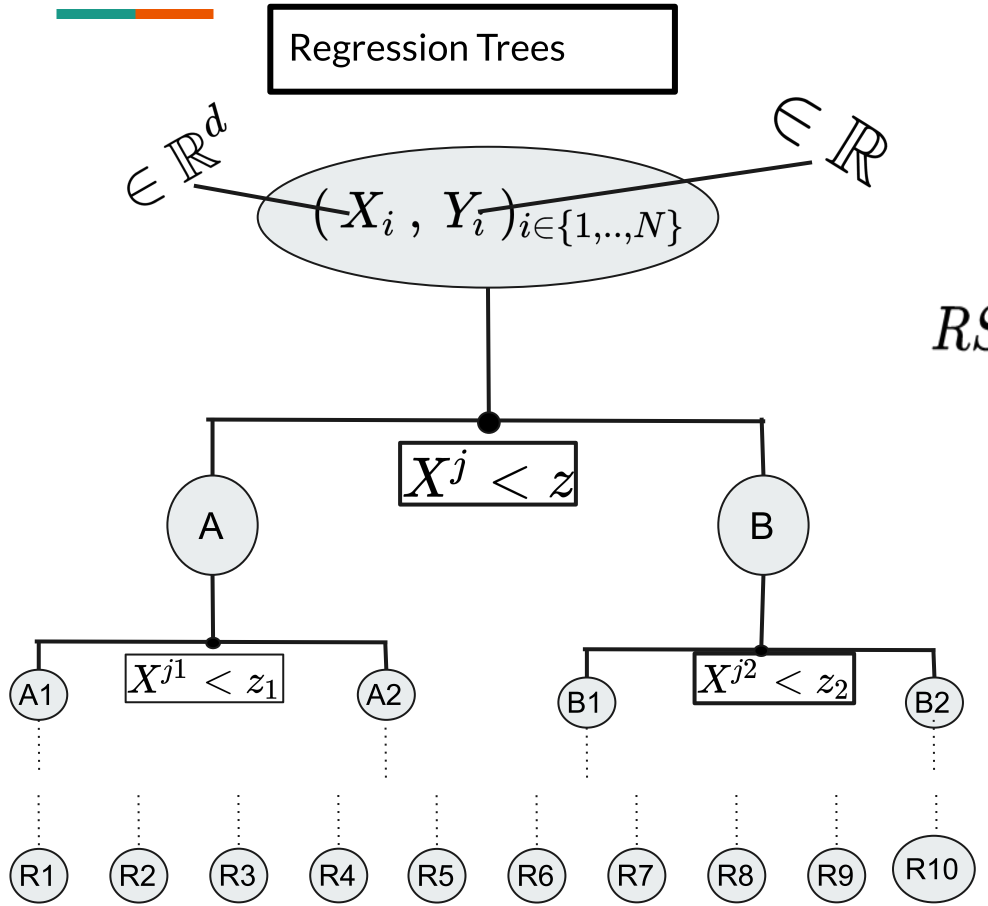
XGBOOST

Regression Trees



$$RSS = \sum_{i \in A} (y_i - \bar{y}_A) + \sum_{i \in B} (y_i - \bar{y}_B)$$

XGBOOST



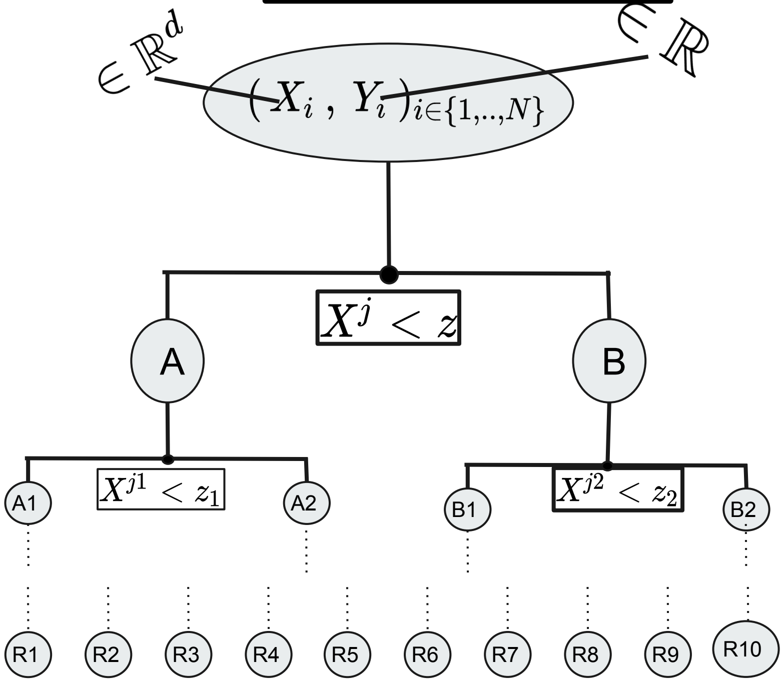
$$RSS = \sum_{i\in A}(y_i - \bar{y}_A) + \sum_{i\in B}(y_i - \bar{y}_B)$$

New X* :
Goes through the tree,
Ends up in a region R

$$\widehat{Y^*} = \bar{Y_R}$$

XGBOOST

Regression Trees



Boosting

Idea : Boost the performances of a weak model by giving a stronger weight to wrongly predicted examples.

- Take randomly 65 % of the observations
- Train a first tree
- Make prediction for the 25% others
- Give a strong weight to the wrong predicted examples.

At the end I trained M models and the final prediction is the mean of each prediction.

DATA

- **X : GCM outputs on a chosen domain**
    - 4 altitude fields : ZG, TA, HUS, (UA,VA) at 850, 700 and 500 hPa
    - 3 surface variables : TAS, PR, (UAS,VAS)
    - on the red domain centered in Montpellier ([-5,10]E x [35,50]N)

  ⇒ for each of them we perform PCA and keep 20 components
    - GHG
    - 2 cosinus and sinus vectors to control the seasonality $cos(\frac{2*\pi*day}{365})$

  ⇒ X_t of size ~150

- **Y : RCM output ⇒ Surface Temperature at the grid point of Montpellier.**

Centered mode : $\boxed{Y = TAS_{RCM,Mpl} - TAS_{GCM,Mpl}}$

Validation in Perfect Model framework

- Common approach to evaluate statistical downscaling methods:

Received: 2 October 2017 | Revised: 13 January 2018 | Accepted: 17 January 2018 | Published on: 23 March 2018
DOI: 10.1002/joc.5462

International Journal of Climatology RMetS

VALUE SPECIAL ISSUE ARTICLE

An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment

J. M. Gutiérrez[1] | D. Maraun[2] | M. Widmann[3] | R. Huth[4,14] | E. Hertig[5] | R. Benestad[6] | O. Roessler[7] | J. Wibig[8] | R. Wilcke[9] | S. Kotlarski[10] | D. San Martín[1,11] | S. Herrera[12] | J. Bedia[1] | A. Casanueva[12] | R. Manzanas[1] | M. Iturbide[1] | M. Vrac[13] | M. Dubrovsky[14,22] | J. Ribalaygua[15] | J. Pórtoles[15] | O. Räty[16] | J. Räisänen[16] | B. Hingray[17] | D. Raynaud[17] | M. J. Casado[18] | P. Ramos[18] | T. Zerenner[19] | M. Turco[20] | T. Bosshard[21] | P. Štěpánek[22] | J. Bartholy[23] | R. Pongracz[23] | D. E. Keller[10,24] | A. M. Fischer[10] | R. M. Cardoso[25] | P. M. M. Soares[25] | B. Czernecki[26] | C. Pagé[27]

Transferability in the future climate of a statistical downscaling method for precipitation in France

G. Dayon[1], J. Boé[1], and E. Martin[2]

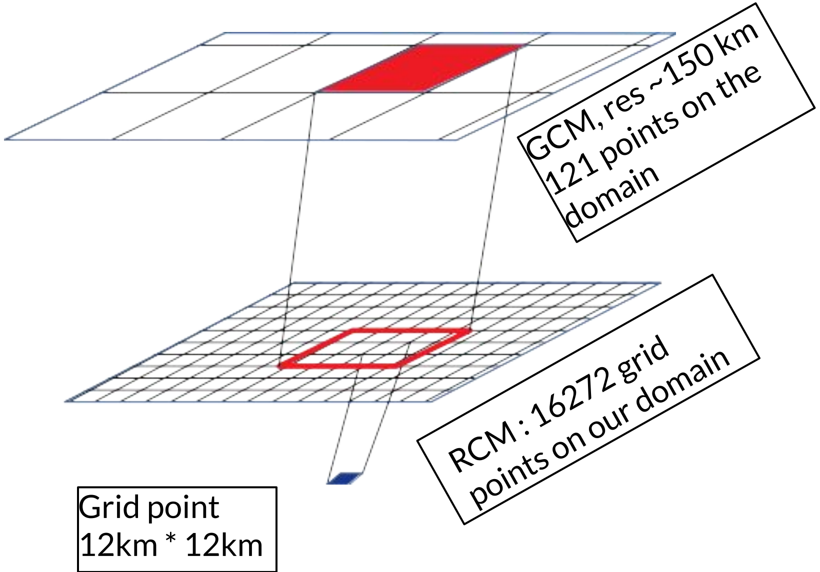
SOME PITFALLS IN STATISTICAL DOWNSCALING OF FUTURE CLIMATE

John R. Lanzante, Keith W. Dixon, Mary Jo Nath, Carolyn E. Whitlock, and Dennis Adams-Smith

Validation in Perfect Model framework

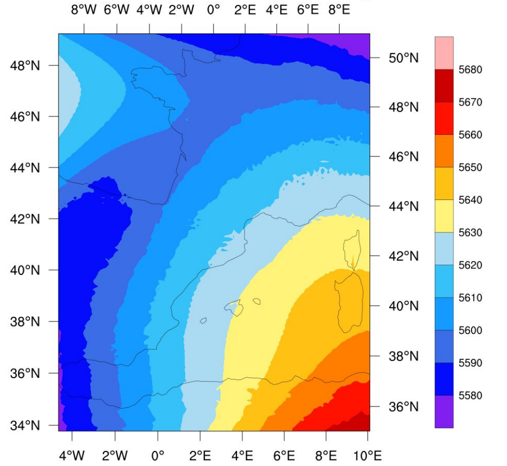
- Common approach to evaluate a statistical downscaling approach :
  - VALUE project (Maraun et. al 2015),
  - Lanzante et al. 2018.
  - Dayon et al. 2015.

- X from a ‘upscaled’ RCM to avoid RCM - GCM chronology mismatch.



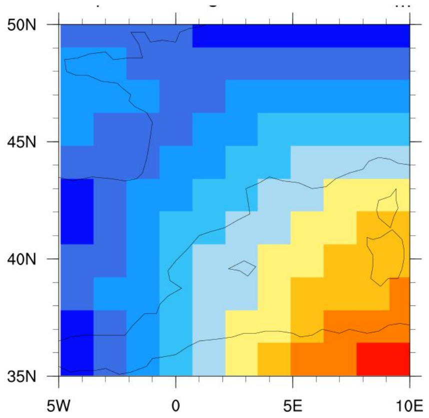
Z500 from the RCM

1/01/2006



Z500 from the upscaled RCM

1/01/2006



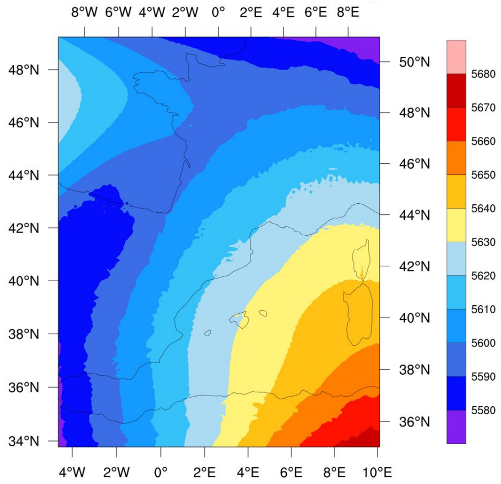
Validation in Perfect Model framework

- Common approach to evaluate a statistical downscaling approach :
  - VALUE project (Maraun et. al 2015),
  - Lanzante et al. 2018.
  - Dayon et al. 2015.

- X from a ‘upscaled’ RCM to avoid RCM - GCM chronology mismatch.

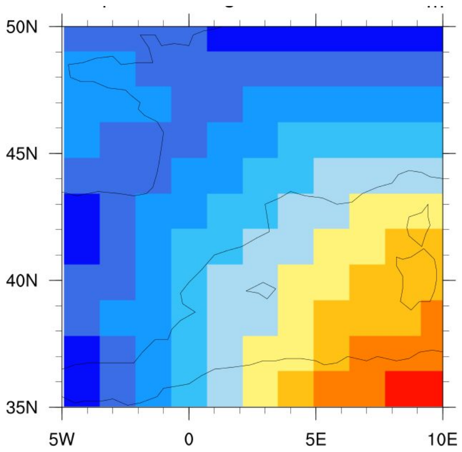
Z500 from the RCM

1/01/2006



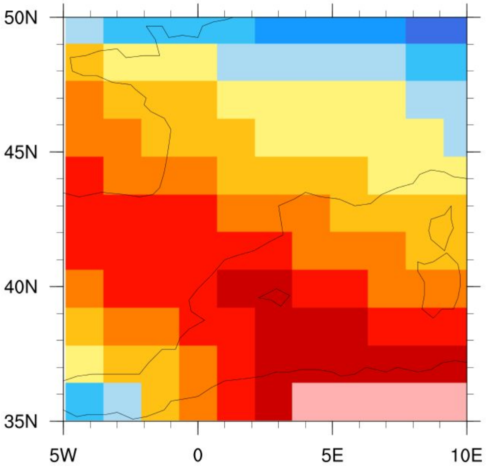
Z500 from the upscaled RCM

1/01/2006



Z500 from the GCM

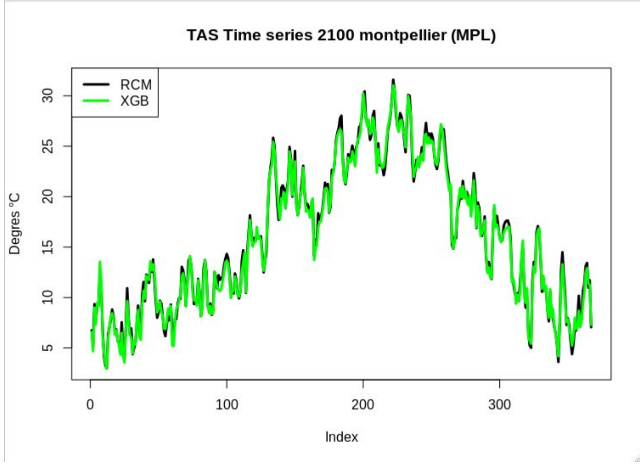
1/01/2006

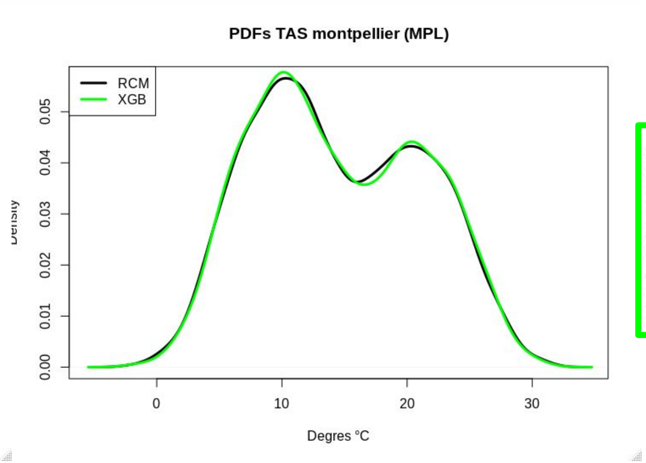


Validation in Perfect Model framework

- Common approach to evaluate a statistical downscaling approach : VALUE project (Maraun et. al 2015), Lanzante, Dixon, Nath 2018.
- X from a 'upscaled' Aladin to avoid chronology distorsion between RCM and GCM.

- **ALADIN 12 km** simulation, forced by CNRM - CM5 over the period **2006-2100** with scenario **RCP4.5 .** Training set : 70% of the years, Testing set: 30% of the years





		DJF	MAM	JJA	SON	ANN
Rmse	(°C)	0.591	0.515	0.648	0.539	0.575
Cor		0.980	0.976	0.968	0.979	0.976
bias	(°C)	0.018	0.011	0.058	-0.017	0.018
pdf		0.039	0.030	0.047	0.028	0.020
extrems		0.126	0.140	0.142	0.273	0.106
extrems bas		0.222	0.099	0.141	0.082	0.137

Validation in Perfect Model framework

- Common approach to evaluate a statistical downscaling approach : VALUE project (Maraun et. al 2015), Lanzante, Dixon, Nath 2018.
- X from a 'upscaled' Aladin to avoid chronology distorsion between RCM and GCM.

- **ALADIN 12 km** simulation, forced by CNRM - CM5 over the period **2006-2100** with scenario **RCP4.5 .** Training set : 70% of the data , Testing set: 30% of the data

Conclusions :

❖ **XGBOOST works well as a Statistical Downscaling method,**

❖ **Able to learn a non-stationary relationship.**

Application : Downscaling of real GCM

2 approaches / philosophies :

X from the upscaled RCM :

- Use the model train in the perfect model
- Focus only on the downscaling action of the RCM
- A too perfect relationship ?

X from the GCM :

- More intuitive approach : learn on a pair (GCM, RCM)
- ‘Learn’ the chronology modification
- More difficult to apply to an other GCM ?

Application : Downscaling of real GCM

2 approaches / philosophies :

X from the upscaled RCM :

- Use the model train in the perfect model
- Focus only on the downscaling action of the RCM
- A too perfect relationship ?

Only the way to train the models changes

- Same variables in X
- Same Y :

$$TAS_{RCM,Mpl} - TAS_{GCM,Mpl}$$

Application to new GCM :
We give the same data to both models.

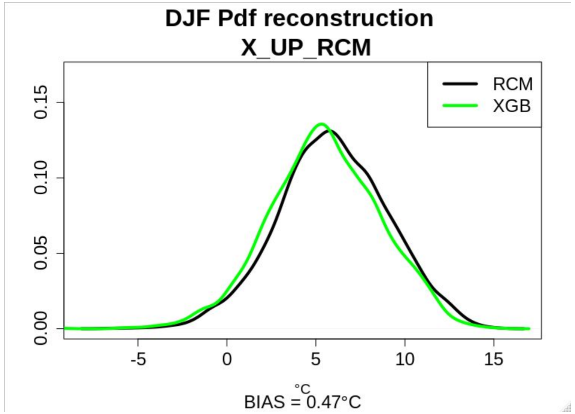
Training sample : RCP 4.5 , 2006-2100

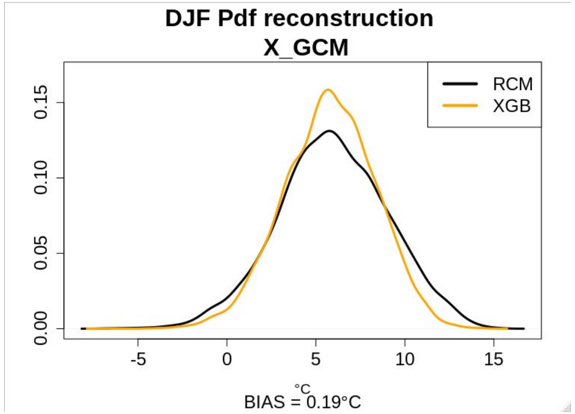
X from the GCM :

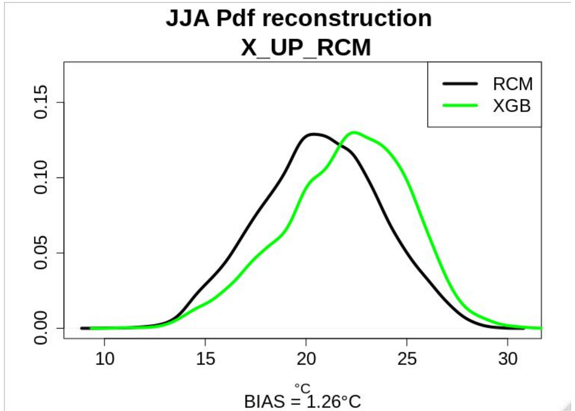
- More intuitive approach : learn on a pair (GCM, RCM)
- 'Learn' the chronology modification
- More difficult to apply to an other GCM ?

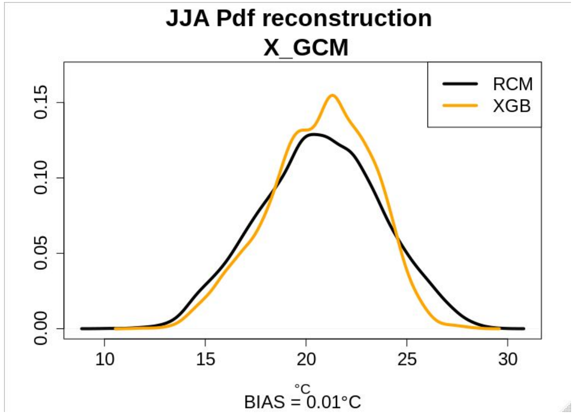
Application : Downscaling of real GCM

HISTORICAL : 1951 - 2005





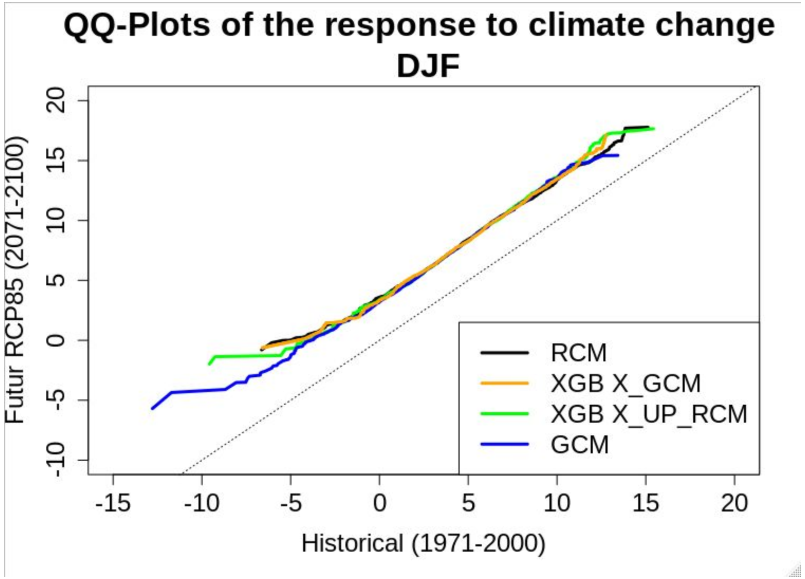


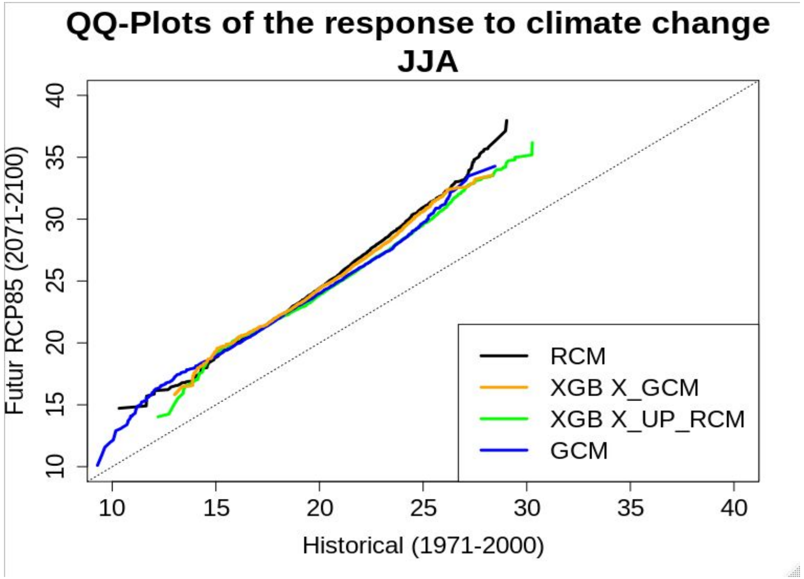


Application : CNRM-CM5 , RCP 8.5

Climate change response 2071-2100 vs 1970-2000

	Annual	DJF	JJA
RCM	3.9 °C	3.4 °C	4.7 °C
X UP_RCM	3.7 °C	3.4 °C	4.2 °C
X GCM	3.9 °C	3.4 °C	4.5 °C
GCM	3.6 °C	3.3 °C	4.2 °C





Conclusion & Perspective

- Building a statistical emulator of a RCM seems feasible
  - Good reproduction of the mean state
  - Some problems with the extremes , warm extremes specially

- Should be tested with other GCMs

- Need to think about multi-variable and 2 Dimensions

- Construction of different models :
  - Neural Network : CNN, LSTM
  - A more complex approach based on a better decomposition of the signal

- Application to CPMs.

Conclusion & Perspective

- Building a statistical emulator of a RCM seems feasible
  - Good reproduction of the mean state
  - Some problems with the extremes , warm extremes specially

- Should be tested with other GCMs

- Need to think about multi-variable and 2 Dimensions

- Construction of different models :
  - Neural Network : CNN, LSTM
  - A more complex approach based on a better decomposition of the signal

- Application to CPMs.
  - ➢ If you are running simulations, it would be nice to save some altitude variables.